

CLAIMS

1. An electronic image capture apparatus comprising: an image
5 detecting device adapted to capture a set of sub-images or tiles
corresponding to different areas of a document at known locations and
electronic processing means adapted to receive the set of sub-images
produced by the device and to process the sub-images to form a machine-
readable text document equivalent to the portion of the document covered
10 by the set of sub-images; characterised in that:

the processing means includes an optical character recognition sub-
routine which is adapted to produce a first set of processable data files
which each comprise a data set of characters corresponding to characters
appearing in a respective sub-image in the set and the relative location of
15 the characters in that sub-image; and

the processing means is adapted to stitch together the characters
stored in the data files to produce a machine readable text document.

2. An electronic image capture apparatus according to claim 1 wherein
20 the image detecting device comprises an electronic camera having a
detector, a lens having a field of view which is adapted to limit the
radiation incident upon the detector to that within the field of view, an
actuator for moving the field of view of the camera relative to the
document to be imaged, and a controller for controlling the actuator to
25 move the field of view of the camera across the document so as to capture
the set of sub-images or tiles.

3. An electronic imaging apparatus according to claim 1 or claim 2
wherein the data in the first set of processable data files is stitched
30 together to produce the machine readable document by allocating

characters in the data files onto corresponding locations in a spatial template (or map) of the machine readable document.

4. An electronic imaging apparatus according to claim 1, 2 or claim 3
- 5 wherein the processing means establishes a co-ordinate system which defines the template of the machine readable document whereby any point in the imaged document can be uniquely identified by its co-ordinate in the machine readable text document.
- 10 5. An electronic imaging apparatus according to claim 4 wherein a second co-ordinate system is defined for each sub-image and the characters located in each sub-image after OCR are stored in the processable data files along with their location in this second co-ordinate frame.
- 15 6. An electronic imaging apparatus according to claim 5 wherein the first and second co-ordinate systems are the same or are related through a transform whereby each character stored in a processable data file can be mapped onto the co-ordinate frame of the machine readable document.
- 20 7. An electronic imaging apparatus according to any preceding claim wherein the sub-images overlap spatially at least by the width of the largest character which is expected in the document.
- 25 8. An electronic imaging apparatus according to any preceding claim wherein where only one data file contains a character at a given location in the machine readable text document the processing means is adapted to allocate that character to that location.

9. An electronic imaging apparatus according to claim 8 wherein if none of the processable data files contain a character for a location in the machine-readable text document then a space is entered in the text document at that location.

5

10. An electronic imaging apparatus according to claim 8 or claim 9 wherein the processing means is adapted to determine the reliability of the data in the processable data files.

10 11. An electronic imaging apparatus according to claim 10 wherein in the event that two or more data files contain different characters corresponding to the same location in the machine readable text file the processing means is adapted to select which data to allocate based on the reliability of the data.

15

12. An electronic imaging apparatus according to claim 10 or claim 11 wherein the processing means is adapted to determine the reliability of the data by applying one or more logical rules to the data in the processable data files.

20

13. An electronic imaging apparatus according to claim 12 wherein the logical rules include using the character which is located furthest away from the edge of a sub-image to construct the machine readable document if there is a conflict.

25

14. An electronic imaging apparatus according to any claim preceding wherein the processing means is adapted to identify lines of text within each processable data file from the spatial distribution of the characters identified within each sub-image.

30

15. An electronic imaging apparatus according to any preceding claim wherein the data in the processable data files comprises ASCII characters.

16. In combination, a document carrying text comprising a combination of characters from a character set and an imaging apparatus according to any one of claims 1 to 15, and in which the apparatus is adapted to obtain sub-images which overlap spatially by a distance greater than the size of at least some of the characters on the document that are present in the region of spatial overlap.

17. A method of creating a machine readable text document in a memory comprising:

capturing an image of a document being scanned by capturing a plurality of sub-images or tiles which correspond to known regions of a document and which in combination cover the document being scanned;

performing an optical character recognition process on each sub-image or tile to create a plurality of text records with machine-readable coded representations of recognised characters; and

joining the text records corresponding to the aligned sub-images tiles so as to create the machine readable text document.

18. The method of claim 17 in which the text records comprising coded representations of recognisable characters are joined by comparing their data content at regions of expected overlap.

19. The method of claim 16 or claim 18 in which the joining operation comprises allocating characters in the text records corresponding to known regions of the document to a corresponding region of the machine readable text document.

20. The method of claim 19 wherein in the event that more than one text record contains a character for the same region of the text document, as occurs at overlap areas, then logical rules are applied to select which character to allocate to that region.

5

21. A computer readable medium having a program recorded therein in which the program causes, in use, a computer running the program to execute the method of claims 17, 18, 19 or 20 or produce an apparatus in accordance with any one of claims 1 to 16.

10

22. A software carrier carrying image processing software which when operational on a computer or network which is connected to a camera either provides the apparatus of any one of claims 1 to 16 or operates the computer or network in accordance with any one of claims 17 to 20.